The Kolmogorov-Smirnov test and its use for the identification of fireball fragmentation

Ivan Melo[a], Boris Tomášik[b,c], Giorgio Torrieri[d], Sascha Vogel[e],

Marcus Bleicher[e], Samuel Koróny[b], Mikuláš Gintner[a,b]

[a] *Žilinská Univerzita, Univerzitná 1, 01026 Žilina, Slovakia*
[b] *Univerzita Mateja Bela, Tajovského 40, 97401 Banská Bystrica, Slovakia*
[c] *Faculty of Nuclear Science and Physics Engineering,*
Czech Technical University in Prague, Břehová 11, 11519 Prague, Czech Republic
[d] *Frankfurt Institute of Advanced Studies, Johann Wolfgang Goethe Universität,*
Ruth-Moufang-Str. 1, 60438 Frankfurt am Main, Germany
[e] *Institut für theoretische Physik, Johann Wolfgang Goethe Universität,*
Max-von-Laue-Str. 1, 60438 Frankfurt am Main, Germany
(Dated: February 10, 2009)

We propose an application of the Kolmogorov-Smirnov test for rapidity distributions of individual events in ultrarelativistic heavy ion collisions. The test is particularly suitable to recognise non-statistical differences between the events. Thus when applied to a narrow centrality class it could indicate differences between events which would not be expected if all events evolve according to the same scenario. In particular, as an example we assume here a possible fragmentation of the fireball into smaller pieces at the quark/hadron phase transition. Quantitative studies are performed with a Monte Carlo model capable of simulating such a distribution of hadrons. We conclude that the Kolmogorov-Smirnov test is a very powerful tool for the identification of the fragmentation process.

I. INTRODUCTION

The highly excited matter created in ultrarelativistic nuclear collisions expands very fast. It is commonly accepted that a deconfined phase has been reached in Au+Au collisions at RHIC [1, 2, 3, 4], while the onset of deconfinement has been advocated at SPS energies [5, 6].

While in lattice QCD calculations a static thermodynamic medium is assumed, in heavy ion collisions the situation is vastly different. Here, the longitudinal expansion dynamics leads to a rapid passage from the deconfined to the confined phase. A system which undergoes the phase transition quickly may not follow the usual equilibrium scenario. In fact, for a first-order phase transition, the high temperature phase may survive down to temperatures drastically below the transition temperature, i.e. the system supercools. If the expansion rate is faster than the nucleation rate of bubbles of the new phase, the system reaches the point of spinodal instability[1]. Beyond such a point, entropy is gained if the system separates into two phases and so it becomes mechanically unstable. Spinodal fragmentation connected with nuclear liquid/gas phase transition has been identified in heavy ion collisions at few hundred MeV per nucleon [7, 8], and it has been proposed that it might be the actual scenario at ultrarelativistic energies as well [9, 10]. Fragmentation assumes that at the phase transition the system decays into droplets of smaller size. These droplets then emit hadrons.

Lattice calculations indicate, however, that at RHIC and LHC the transition from partonic to hadronic matter is a rapid but smooth crossover [11]. Thus spinodal decomposition seems irrelevant scenario in this case. On the other hand, conformal symmetry is broken close to the phase transition and as a consequence the *bulk* viscosity—being negligible otherwise—shows a peak here [12, 13, 14]. Bulk viscosity acts against the expansion and slows it down. As a result, if the system previously accumulated kinetic energy due to expansion, it may fragment [14]. An analysis of hydrodynamic instabilities shows that such a scenario may be realistic [15].

Hence, it appears that fragmentation may happen in ultrarelativistic nuclear collisions. Many kinds of observables might be sensitive to it. Most notable are multiplicity fluctuations in varying rapidity windows [16], fluctuations of mean p_t [17], rapidity correlations [18, 19], proton and kaon correlations [18], ϕ-meson production [18], and two-pion femtoscopy [14, 20, 21].

In this paper we inspect event-by-event fluctuations of rapidity distributions. If final state hadrons are emitted from droplets, their velocities will be close to those of the droplets. Thus, clustering would appear in their momentum distribution. Moreover, in each event clusters will have different velocities. An important contribution to clustering will also come from the resonance decays and we shall investigate this effect. Thus, the momentum distribution will vary from event to event. Specifically, here we shall compare *rapidity distributions* from different events and look for differences due to fireball fragmentation. To this end, we employ the Kolmogorov-Smirnov (KS) test which can be used for identification

---

[1] This is the inflection point of the dependence of entropy on an extensive variable, see e.g. [7, 8].
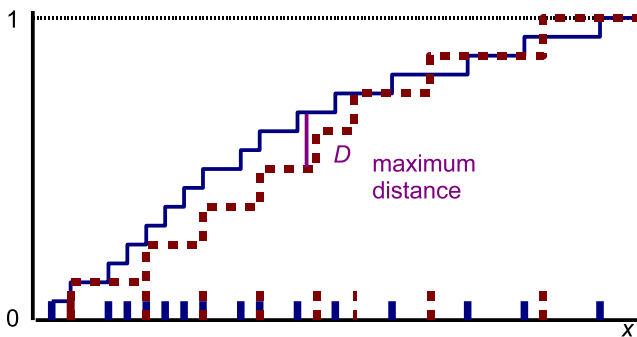
FIG. 1: Construction of the two empirical cummulative distribution functions, one with thin solid (blue) line and one with thick dashed (red) line. The maximum distance between them is D.

of non-statistical differences between two empirical distributions [22, 23]. An important advantage of the KS test is its independence from the underlying distribution of the measured quantity.

In the next Section we shortly introduce the KS test. Then, we illustrate its sensitivity by a couple of toy simulations. For more realistic studies we generate artificial data with the help of the event generator DRAGON [24] which is very briefly introduced in Section IV. Results obtained with these data are presented in Section V. We conclude in Section VII. In the Appendix we review the evaluation of the cummulative distribution function for the Kolmogorov distribution.

II. THE KOLMOGOROV-SMIRNOV TEST AND HOW TO USE IT

Let us start by explaining the technical part of the problem. One has two empirical distributions in variable x, which can be rapidity, p_t, or yet something else[2]. (In the present work we work with rapidities.) The multiplicities may differ. The question we want to ask is, whether the two empirical distributions are the same in the sense that they would correspond to the same underlying theoretical single-particle probability density, and there are no correlations between particles in one event.

Practically, the quantity x is measured for each particle in an event. The empirical cummulative distribution function (ECDF) is constructed so that a step of the height $1/n_i$ (n_i is the multiplicity of the event) is made on all positions of measured x's (Fig. 1). This is done for both events of a pair. Subsequently, one finds the *maximum* vertical distance between the two ECDF's

---

[2] Note that for a cyclic variable, e.g. azimuthal angle ϕ, the KS test cannot be applied and instead a modification known as Kuiper test [25, 26] must be employed.

and introduces

$$d = \sqrt{n}D = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} D \qquad (1)$$

where D is the distance of two ECDF's and n_1, n_2 are the multiplicities of the two data sets. The procedure of the test is illustrated in Figure 1. The cummulative distribution function of the *Kolmogorov distribution* concerns the case of events generated from the same underlying theoretical probability distribution for the quantity x, $\rho(x)$. It will be expressed with the help of the function $Q(d)$ as

$$P(d' < d) = 1 - Q(d). \qquad (2)$$

where $P(d' < d)$ is the probability that we find a difference d' smaller than d. An important feature of this approach is that $Q(d)$ does not depend on the particular shape of the theoretical distribution $\rho(x)$. Unfortunately, the general form of $Q(d)$ valid for any multiplicities and distances d is not suitable for practical evaluation. Usable approximate expressions for $Q(d)$ are summarised in the Appendix. It follows that if many pairs of events would be drawn from a set of events generated all from the same underlying probablility distribution and for each pair the quantities d and $Q(d)$ would be determined, then the $Q's$ would be distributed *uniformly*. Deviation from uniform distribution, particularly an enhanced population of low Q's (large d's) indicates that the events are *not drawn independently from the same underlying distribution*. In this way the KS test will be used here.

Note that the KS test does not identify the physical origin of the difference between the events. It is a robust way to identify that there is a difference, however, the origin must be singled out by other means. In addition to fireball fragmentation, these can be fluctuations of the initial state of the fireball evolution, final resonance decays, conservation laws, and quantum correlations. A detailed investigation of these will be pursued in subsequent papers. The important message of the KS test is, that it can disprove the usual paradigm that data from many collisions (within the same centrality class) are produced by basically identical fireballs.

It is important to realise that the number of significant decimal figures to which the quantity x (rapidity here) is measured may also influence the result of the KS test if it is applied on a large number of pairs of events. This is illustrated in Figure 2. The peak at $Q \to 1$ increases with lowering the number of decimal places taken into account. The explanation is trivial but instructive. Only rapidities between 0 and 1 were generated. Within 10^5 events, each having multiplicity around 200, there are about 2×10^7 particles. Hence, if their values of x are given to less than 8 figures, we are guaranteed to have repeating values of x in our sample. For 6 given figures we expect each value to appear on average 20 times, for 4 figures it is 2000 times and for 2 figures we are even at 200,000 times! Clearly, this effect correlates the events since it artificially chooses
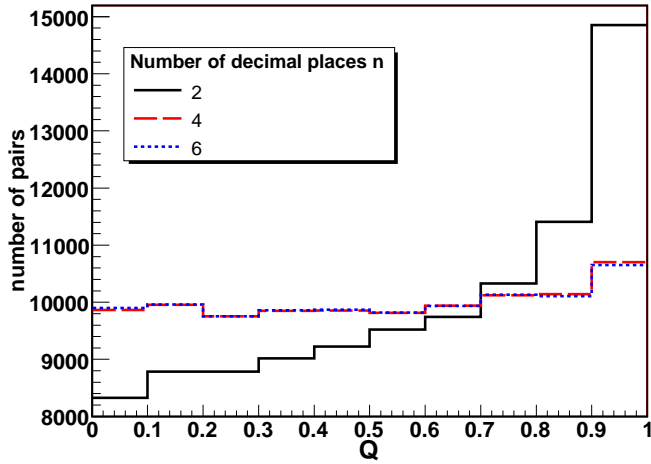
FIG. 2: Histograms of Q's from the KS test applied on 10^5 pairs out of 10^5 events generated from a uniform distribution between 0 and 1 in the variable x (rapidity) and with multiplicities distributed according to Poisson distribution with the mean 200. Different histograms correspond to rapidity data truncated after 2, 4, and 6 decimal places. The histogram with 8 significant figures is identical to that with 6 figures.

x from a finite number of possible values. According to its construction, the D's will acquire smaller values on average.

We have also checked that the normal fluctuation of the number of entries in a bin of the Q-histogram is equal to the square root of the number of entries. We thus propose the use of quantity

$$R = \frac{N_0 - \frac{N_{tot}}{B}}{\sigma_0} = \frac{N_0 - \frac{N_{tot}}{B}}{\sqrt{\frac{N_{tot}}{B}}} , \qquad (3)$$

where N_0 is the number of pairs in the first bin of the Q-histogram (next to $Q = 0$), N_{tot} is the total number of pairs, B is the number of bins of the Q-histogram, and $\sigma_0 = \sqrt{N_{tot}/B}$ is the expected variance of the number of entries in the first bin. The modulus of R should be of the order 1; values considerably bigger than that indicate non-statistical differences between the events.

III. THE SENSITIVITY OF THE TEST

In this section we study how the proposed method works in case of clear cut examples. First, we generate samples of "events" where one half of all events is generated according to Gaussian distribution with the width $\sigma = 0.1$. For the second half of events we keep the same width and vary the mean: we have samples with the mean shifted with respect to the other half by 2σ, 1σ, 0.5σ, 0.1σ, and 0.01σ. In Figure 3 we observe how the KS procedure recognises the difference of 0.1σ pretty well if the average multiplicity is 512, and how the resolution power decreases when lowering the multiplicity
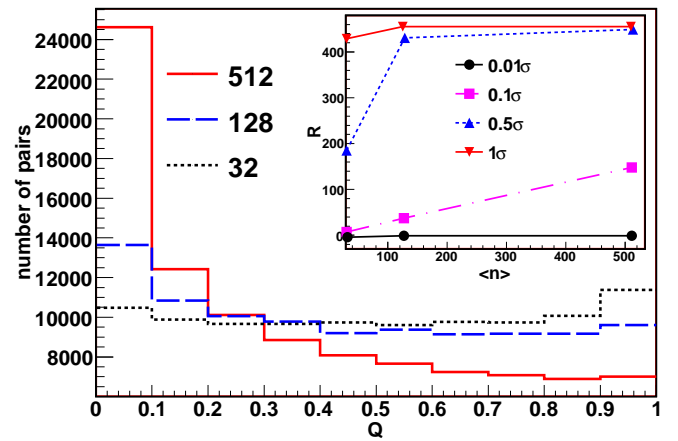

FIG. 3: (Color online) Large plot: the Q histograms from the KS test on event samples consisting from two classes of events. One class was generated from Gaussian distribution with the mean 0 and the width $\sigma = 0.1$. The other class is generated from Gaussian distribution with the same width, but the mean is shifted by 0.1σ. The multiplicities of the events are 32 (black dotted histogram), 128 (blue dashed), and 512 (red solid). Smaller inset plot: the dependence of the parameter R on the multiplicity of the events for the difference of the means equal to 0.01σ (black circles), 0.1σ (magenta squares), 0.5σ (blue triangles), and 1σ (red upside down triangles).
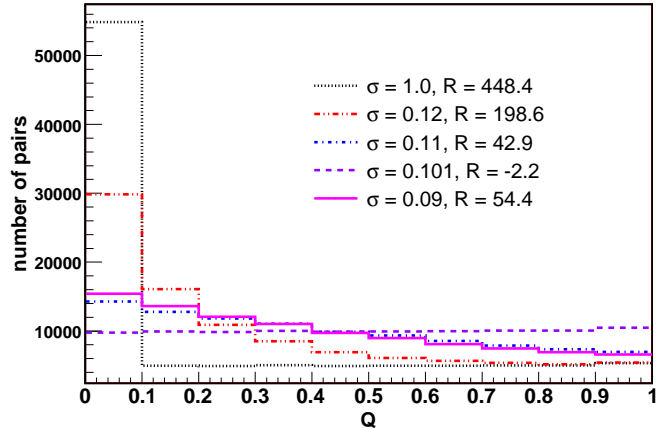

FIG. 4: The Q-histograms from event samples consisting of two classes of events where rapidities were generated from Gaussian profiles with the same mean and the multiplicity was distributed Poissonian with the mean 512. The width in one class of events was 0.1. The width in the second class has been varied; different histograms correspond to different widths. The values of R are written in the legend.

to 32. As seen in the smaller inset plot, for smaller distances between the two Gaussian means the difference is not recognised, while for larger distance the difference is resolved by the test for all multiplicities.

In Figure 4 we explore the effect of a variable width. One half of events was simulated with Gaussian distribu-
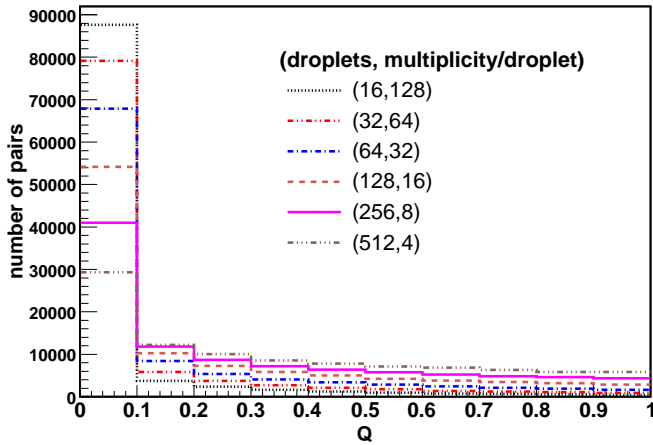
FIG. 5: The effect of the number of Gaussian sources on the Q-histogram. In the legend, left is the number of Gaussian sources distributed uniformly between -1 and 1, right is the average number of pions from each source. The width of each Gaussian source was 0.707.

tion with the width of 0.1 and the other one with the same mean but a different width. The widths are 1.0, 0.12, 0.11, 0.101, and 0.09. The multiplicity was Poissonian-distributed with the mean of 512. We observe that except for the cases where the widths differ by ten per cent or less the difference is picked up by the procedure.

Finally, we test a case which is closest to the fireball fragmentation scenario that we want to explore in detail. In Figure 5 we show the results from a simulation, where each event consists from superposition of many Gaussian distributions. The width of all these distributions is 0.707 and is motivated by the typical rapidity spread of the pion rapidity at a realistic freeze-out temperature. The means of the Gaussians are generated from a uniform distribution between –1 and 1. We test cases with 16, 32, 64, 128, 256, and 512 Gaussians per event, which emit on average 128, 64, 32, 16, 8, 4 particles per Gaussian, respectively, so that the total multiplicity is always 2048. We observe, that even in the least favorable simulation with a large number of small droplets, the difference between events is clearly visible.

IV. MONTE-CARLO DROPLET GENERATOR

Realistic events samples on which the KS test are applied were generated with the help of the Monte Carlo event genarator DRAGON [24]. Here we provide very brief overview of its capabilities.

DRAGON assumes that the fireball decays into droplets which are distributed according to the blast-wave model. Thus their distribution in position and velocity is given by

$$S_D(x, v) \propto H(\eta)\, \Theta(R - r)\, \delta(\tau - \tau_0)\, \delta^{(4)}(v - u(x))\,, \quad (4)$$

where we use polar coordinates r and ϕ, the space-time rapidity and longitudinal proper time

$$\eta = \frac{1}{2} \ln \frac{t + z}{t - z} \quad (5)$$

$$\tau = \sqrt{t^2 - z^2}\,, \quad (6)$$

as coordinates in the space-time. The fireball has a transverse radius R and τ_0 is the Bjorken proper time of the decay. The four-velocity of the droplet v is given by the local flow velocity at the position where the droplet is created,

$$u_\mu(x) = (\cosh \eta \, \cosh \eta_t,\, \cos \phi \, \sinh \eta_t,$$
$$\sin \phi \, \sinh \eta_t,\, \sinh \eta \, \cosh \eta_t)\,, \quad (7)$$

with

$$\eta_t = \frac{\sqrt{2}\rho_0 r}{R}\,, \quad (8)$$

where ρ_0 is a model parameter. (The model is designed so that it can simulate azimuthally non-symmetric fireballs, but we do not explore such a possibility here.) The function $H(\eta)$ specifies the space-time rapidity distribution. It can be uniform or Gaussian. For the present investigation we use the uniform distribution in rapidity.

The volumes of the droplets are random according to a gamma distribution

$$\mathcal{P}_2(V) = \frac{1}{b\Gamma(2)} \frac{V}{b} \exp\left(-V/b\right)\,, \quad (9)$$

with a model parameter b. The droplets decay into hadrons exponentially in time, so the times of emission of the droplets are distributed in the rest frame of the emitting droplet according to $\exp(-\tau/R_D)$, where R_D is the radius of the droplet. A droplet emits hadrons according to thermal distribution with a temperature T_k, until it uses up all of its mass. The mass of the droplet is determined according to its volume and the energy density which is set to $0.7\,\mathrm{GeV\,fm^{-3}}$.

Hadrons may be emitted from the droplets or produced in the remaining space between them. The relative abundance of those emitted from droplets is specified as a model parameter. Hadrons emitted from the bulk are generated according to the blast-wave emission function [27, 28, 29]

$$S(x, p)\, d^4x = \frac{2s + 1}{(2\pi)^3}\, m_t\, \cosh(y - \eta)\, \exp\left(-\frac{p^\mu u_\mu}{T_k}\right)$$
$$\times \Theta(R - r)\, H(\eta)\, \delta(\tau - \tau_0) d\tau\, \tau\, d\eta\, r\, dr\, d\phi\,. \quad (10)$$

Here the factor $(2s + 1)$ denotes spin degeneracy.

Resonances are included in the simulation. They decay according to the standard two-body or three-body kinematics. Probabilities of production of individual species are given by the statistical model with a chemical freeze-out temperature T_{ch} and chemical potentials for baryon number and strangeness.

V. FLUCTUATING RAPIDITY DISTRIBUTIONS

The Monte Carlo event generator DRAGON is employed [24] to simulate realistic data on which the KS test is performed. We use the test on data generated for RHIC Au+Au collisions at $\sqrt{s} = 200\,A\,$GeV and FAIR Au+Au at $\sqrt{s} = 7.6\,A\,$GeV. For the data analysis we accept hadrons within the rapidity interval [–0.5,0.5].

For RHIC, we have generated events with uniform rapidity distribution in the interval [–3,3]. The total hadron multiplicity was set to $dN/dy = 1000$. The chemical composition is determined by the following choice of parameters: $T_{ch} = 155$ MeV, $\mu_B = 26$ MeV [30]. We neglect the strangeness chemical potential. The list of resonances includes mesons up to a mass of 1.5 GeV/c^2 and baryons up to 2 GeV/c^2. The geometry of the decaying fireball is given by the radius $R = 10$ fm and $\tau_0 = 9$ fm/c. The dynamical state of the fireball is set by the kinetic freeze-out temperature $T_k = 150$ MeV and the transverse expansion gradient $\eta_f = 0.6$. We set the volume parameter of the droplets b to the value of 10 fm^3. As a first benchmark, complementary samples of 10,000 events are generated: one with all particles being emitted from droplets, the other with all particles being emitted from the bulk fireball.

As a second benchmark test we generate 10,000 events at the FAIR energy of $\sqrt{s} = 7.6\,A$ GeV where no particles are emitted from droplets. In this case the chemical freeze-out parameters are set to the corresponding values $T_{ch} = 140$ MeV, $\mu_B = 375$ MeV, and $\mu_S = -53$ MeV. The kinetic freeze-out temperature is $T_k = 140$ MeV and the transverse expansion of the fireball is characterised by $\eta_f = 0.4$. Here, the rapidity distribution is Gaussian with a width of 0.7 and the total hadron multiplicity is 1,500. The transverse radius of the fireball and its Bjorken lifetime are 9 fm and 8 fm/c, respectively.

In figure 6 we show the difference between the Q-histograms from events with and without droplets. For the RHIC energy, one observes a characteristic enhancement towards small Q values in the case of particle emission from droplets for all investigated particle species except (anti-)protons. (For the setting without droplets (i.e. only bulk emission) this low Q enhancement is strongly suppressed. Quantitatively, this is reflected in a factor of 10 difference of the extracted R values. The RHIC results without droplet formation are also in line with the results obtained at FAIR energies, showing that the KS test does not produce falsly positive results when going to smaller samples with a different rapidity distribution.

Resonance decays also have a clustering effect on the decay products. Therefore, a signal of clustering is also seen in the set of events without droplets. In case of all hadrons, these are mainly ρ's and Δ's. If we limit our analysis to pions only, then there is correlation due to the ρ. To test this hypothesis, one can perform the KS test with protons only, since there is no resonance that would

decay into two baryons. The drawback of using protons only is limited statistics in two ways. Firstly, their total multiplicity is lower, e.g. there are only 10 to 50 protons in the acceptance per RHIC event. Therefore, one observes fewer pairs at small Q and a peak at Q close to 1 in case there are no droplets (see appendix). Secondly, if the droplets are small, protons are a less ideal probe because a droplet may not have enough energy to emit more than one proton and the correlation is gone then. An alternative solution is to use pions of the same charge. These are more abundant than protons and no strong effect of resonances is seen here. Note, however, that we have not included the effect of pair wave function symmetrisation which leads to Bose-Einstein correlations.

Note also, that resonances not only introduce correlations, they can also weaken the correlations due to droplets. Resonance decay products obtain some momentum due to higher mass of the mother resonance. Thus, the velocities of decay products will be more smeared around the velocity of the droplet which emitted the resonance than the velocities of hadrons emitted from the droplet directly.

The influence of the size of the droplets is studied in Figure 7. The parameters of the simulation are kept the same as in the previous case, but the volume parameter b varied to values 5, 10, 20, 50 fm^3. All particles are emitted from droplets. We do the KS test with charged hadrons. As expected from previous analysis (with $b = 10\,$fm^3), a dominant low Q peak emerges for all droplet volumes down to 5 fm^3. From this we conclude that even small size droplets can be detected with the analysis method presented here.

As a final physics benchmark of the KS test we explore the effect of changing droplet fraction of the total multiplicity. A systematic study of how the percentage of hadrons emitted from droplets affects the result is presented in Figure 8. Here, the size of the droplets is fixed to $b = 10\,$fm^3 and only the percentage of hadrons from droplets is varied. Even if only one quarter of all hadrons comes from the droplets and the rest from the gas in between them, the signal is well visible and the KS test can discriminate between the formation and the non-formation of droplets.

VI. RESOLUTION

Finally, we address the experimentally crucial question whether the droplet signal in the Q-histogram stays recognisable if the rapidities are measured with finite resolution. We investigate this issue in Figure 9. Events simulated with DRAGON for a volume parameter $b = 10\,$fm^3 and with 50 per cent of hadrons emitted from droplets are taken and the generated rapidities are smeared by a Gaussian with widths of 0.1, 0.5, and 1, to mimic the finite resolution of experimental measurements. The events with smeared rapidities are then processed with the KS test. We observe a gradual weak-
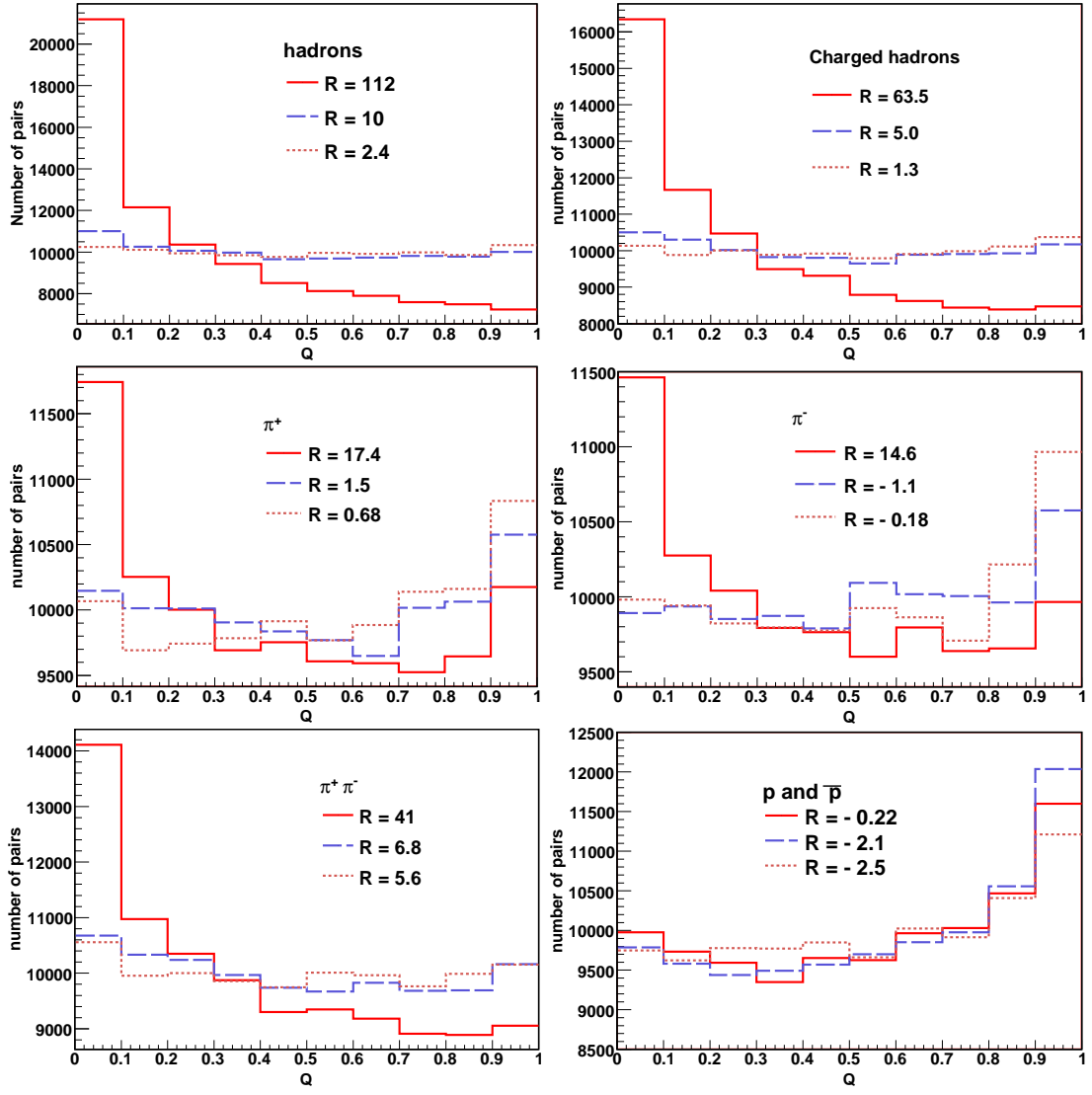
FIG. 6: (Color online) The Q-histograms resulting from simulations of realistic hadronic final states with the help of DRAGON. Solid (red) histograms correspond to simulation of RHIC Au+Au collisions with droplets. Dashed (blue) histograms are from simulations for RHIC without droplets. Dotted (brown) histograms show the results of simulations for nuclear collision at FAIR without fragmentation. Different panels show results obtained for all hadrons, charged hadrons, π^+, π^-, charged pions, protons and antiprotons. The values of R are indicated in the panels.

ening of the signal strength at low Q. For smearing by 0.5 units of rapidity the peak height becomes comparable with the peak resulting from resonance decays only (cf. Figure 6). For even poorer resolution, the peak can not be regarded as an unambiguous signal for droplet formation. The resolution, however, is usually on the level of $\Delta y \approx 0.1$.

VII. CONCLUSIONS

The Kolmogorov-Smirnov test is a powerful tool in searching for non-statistical differences between events. The test itself is more general than investigated here, and

applications will be presented in following papers. The logic of its use is the following: select a class of events which are "as identical as possible", in particular in a very narrow centrality class. Conventional scenarios predict that each event would evolve according to the same scenario and the final distributions of hadrons would be identical in all events. The KS test is able to detect deviations from this scenario. If an effect is observed, it remains to be studied what phenomenon leads to positive results.

As a currently widely discussed topic we focussed the present investigation on the possible decay of the fireball into smaller droplets. The present study showed that the KS test is perfectly suited for this task and allows
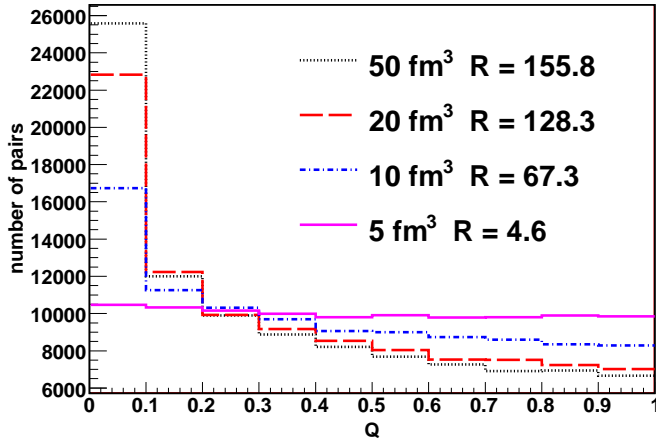
FIG. 7: Q-histograms corresponding to various sizes of droplets: $b = 5, 10, 20, 50 \, \text{fm}^3$. The values of R are shown in the legend. The Q-histograms are obtained from samples of 10,000 events where all hadrons have been emitted from droplets. Only charged hadrons have been taken in constructing these histograms.
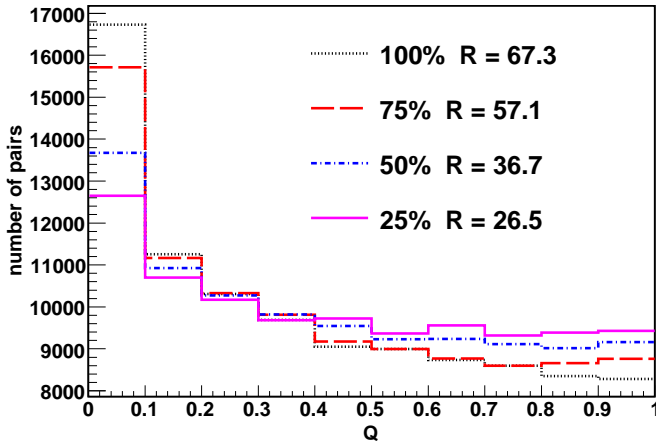

FIG. 9: The influence of finite rapidity resolution. Charged hadrons were generated with the Monte Carlo event genarator DRAGON with $b = 10 \, \text{fm}^3$ and 50% of hadrons generated from droplets. Before the analysis, rapidities were smeared with Gaussian distribution with the width 0.1 (long-dashed line), 0.5 (dash-dotted), 1.0 (short-dashed). Solid line shows the result with non-smeared data.

be performed in subsequent papers.

FIG. 8: Q-histograms obtained from samples of 10,000 events where droplets with a volume parameter of $b = 10 \, \text{fm}^3$ are present. The percentage of hadrons coming from droplets is varied: 25%, 50%, 75%, and 100%. The histograms are constructed with charged hadrons only. The values of R are listed in the legend.

to extract a prominent signal. The signal is robust even if only a small amount of the hadrons come from the droplets and survives realistic final rapidity resolution. Thus, the test can be also used in a negative way: if its application on data yields only limited or no signature of non-statistical event-by-event fluctuations of the rapidity distributions, this puts limits on the scenarios assuming fireball fragmentation.

The investigation of the signal of other effects (including a comparison to full Monte Carlo transport simulation) in the KS test deserves separate studies and shall
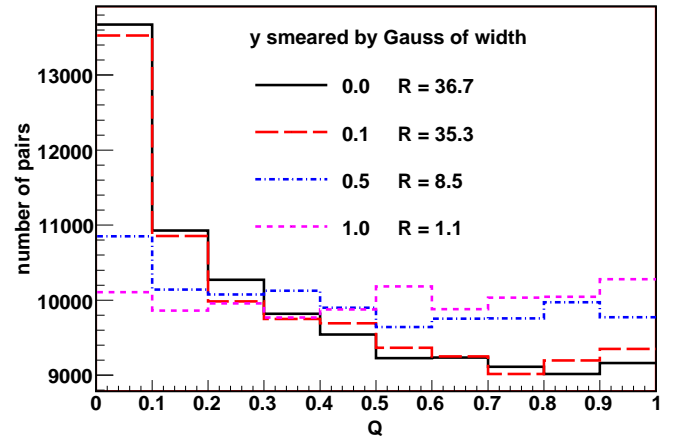
APPENDIX A: EVALUATION OF THE KOLMOGOROV-SMIRNOV DISTRIBUTION

Throughout this paper we use the *two-sample two-sided* (Kolmogorov-)Smirnov test[3]. The cummulative distribution function of the difference D for the one-sample test in case $n \to \infty$ was derived by Kolmogorov

---

[3] The Kolmogorov *one-sample* test refers to a comparison of one empirical cummulative distribution function based on data with a smooth thoretical distribution function. This is distinguished from the two-sample (Smirnov) test where two data samples are compared with each other. One-sided and two-sided tests refer simply to the difference of two cumulative distribution functions or to its absolute value, respectively.

[22]:

$$Q(D) = K_0(D) = -2\sum_{k=1}^{\infty}(-1)^k \exp\left(-2k^2 D^2\right) . \quad \text{(A1)}$$

Later, Smirnov [23] proved that the same result applies for the two-sample test for $n_1, n_2 \to \infty$ under replacement $D \to d = D\sqrt{n_1 n_2/(n_1+n_2)}$. We have checked that such an asymptotic case is not a good approximation even for n's around 200. It is therefore desirable to obtain formulas valid in non-asymptotic case.

A simple solution is to replace the quantity $d = \sqrt{n}D$ in eq. (A1) with the following one, originally due to

Stephens [31] (found also in Numerical Recipes [32]):

$$d = D\left(\sqrt{n} + 0.12 + \frac{0.11}{\sqrt{n}}\right) \quad \text{(A2)}$$

A different formula with a few terms of an expansion in $1/\sqrt{n}$ has been derived by Li-Chien [33, 34]. In such an expansion

$$Q(d) = K_0(d) + K_1(d) + K_2(d) + K_3(d) + \dots . \quad \text{(A3)}$$

The leading order term has been displayed in eq. (A1). The following three terms are

$$K_1(d) = \frac{4d}{3\sqrt{n}}\sum_{k=1}^{\infty}(-1)^k k^2 \exp\left(-2k^2 d^2\right) \quad \text{(A4)}$$

$$K_2(d) = \frac{1}{9n}\sum_{k=1}^{\infty}(-1)^k \left(k^2 - \frac{1}{2}\left(1-(-1)^k\right) - 4k^2 d^2\left(k^2 - \frac{1}{2}\left(1-(-1)^k\right)+3\right) + 8k^4 d^4\right)\exp\left(-2k^2 d^2\right) \quad \text{(A5)}$$

$$K_3(d) = -\frac{2d}{27n^{3/2}}\sum_{k=1}^{\infty}(-1)^k k^2 \left(\left(k^2 + \frac{22}{5} - \frac{3}{2}\left(1-(-1)^k\right)\right)\right.$$
$$\left. - k^2 d^2 \left(\frac{4}{3}k^2 + \frac{88}{15} - 2\left(1-(-1)^k\right)+12\right) + 8k^4 d^4\right)\exp\left(-2k^2 d^2\right) . \quad \text{(A6)}$$

These relations were derived for one-sample test. We checked, however, that they are much easier to handle and give better results when tested on samples of statistically identical events (see below) than approximations to two-sample distributions for non-asymptotic cases [35, 36]. Therefore, we decided to use these relations although we note that a revision of the formulae for two-sample tests is desirable.

In practical calculations it turns out that it is sufficient to cut off the expansions in eqs. (A1), (A4–A6) at $k = 4$. To illustrate this point, we compare in Fig. 10 the histograms of Q's calculated from eq. (A2) with histograms based on eq. (A3) with the cut-off at $k = 2, 3$ and 4 respectively. We show the results for 10^5 pairs of events chosen randomly out of 10^5 simulated events. Each sim-

ulated event is represented by 'rapidities' uniformly generated on (0,1) with Poissonian distributed multiplicities with mean values of 50, 200 and 1000. Thus we calculate $Q(d)$ using the expansion (A3) up to the term K_3 and evaluate the sums in eqs. (A1), (A4–A6) up to the fourth order in k.

The Li-Chien approximation truncated after $k = 4$ may lead to negative Q for $d > 1.94$. Such a pair would fall into the last Q-bin. To fix the problem for this value of d an approximation due to Marsaglia [37] is employed

$$Q(d) = 2\exp\left[\left(-2.000071 - \frac{0.331}{\sqrt{n}} - \frac{1.409}{n}\right)nd^2\right] . \quad \text{(A7)}$$

[1] I. Arsene *et al.* [BRAHMS Collaboration], Nucl. Phys. A **757**, 1 (2005) [arXiv:nucl-ex/0410020].
[2] B. B. Back *et al.*, Nucl. Phys. A **757**, 28 (2005) [arXiv:nucl-ex/0410022].
[3] J. Adams *et al.* [STAR Collaboration], Nucl. Phys. A **757**, 102 (2005) [arXiv:nucl-ex/0501009].
[4] K. Adcox *et al.* [PHENIX Collaboration], Nucl. Phys. A **757**, 184 (2005) [arXiv:nucl-ex/0410003].

[5] C. Alt *et al.* [NA49 Collaboration], Phys. Rev. C **77**, 024903 (2008) [arXiv:0710.0118 [nucl-ex]].
[6] U. W. Heinz and M. Jacob, arXiv:nucl-th/0002042.
[7] P. Chomaz, M. Colonna and J. Randrup, Phys. Rept. **389**, 263 (2004).
[8] J. Randrup, Nucl. Phys. A **752**, 384 (2005).
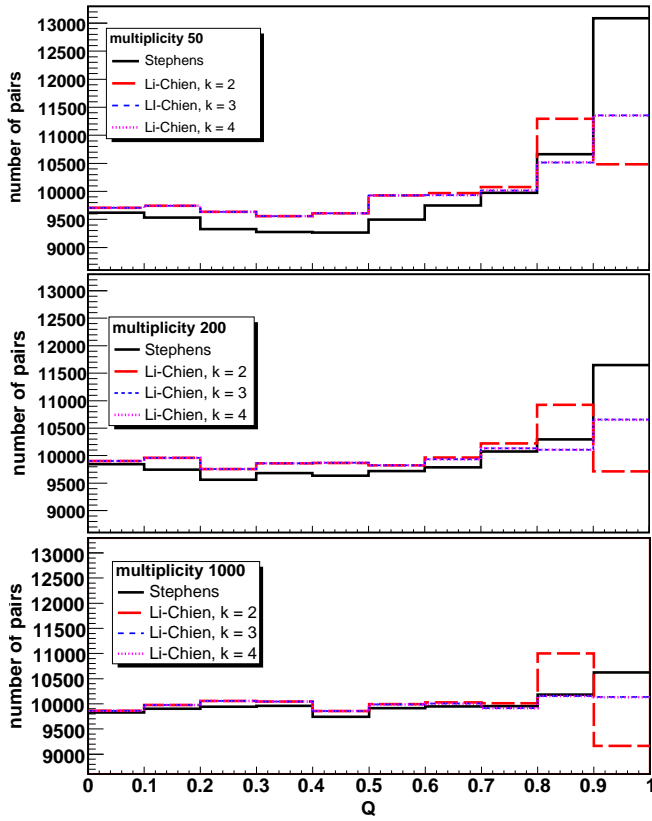[9] I. N. Mishustin, Phys. Rev. Lett. **82**, 4779 (1999) [arXiv:hep-ph/9811307].

FIG. 10: Stephens' approximation according to eq. (A2) vs. Li-Chien approximation (eq. (A3)) with varying number of terms after which it is truncated. The short-dashed blue line coincides with the dotted purple line.

[10] O. Scavenius, A. Dumitru, E. S. Fraga, J. T. Lenaghan and A. D. Jackson, Phys. Rev. D **63**, 116003 (2001) [arXiv:hep-ph/0009171].

[11] Y. Aoki, G. Endrodi, Z. Fodor, S. D. Katz and K. K. Szabo, Nature **443**, 675 (2006) [arXiv:hep-lat/0611014].

[12] K. Paech and S. Pratt, Phys. Rev. C **74**, 014901 (2006) [arXiv:nucl-th/0604008].

[13] D. Kharzeev and K. Tuchin, arXiv:0705.4280 [hep-ph].

[14] G. Torrieri, B. Tomášik and I. Mishustin, Phys. Rev. C **77**, 034903 (2008) [arXiv:0707.4405 [nucl-th]].

[15] G. Torrieri and I. Mishustin, arXiv:0805.0442 [hep-ph].

[16] G. Baym and H. Heiselberg, Phys. Lett. B **469**, 7 (1999) [arXiv:nucl-th/9905022].

[17] W. Broniowski, B. Hiller, W. Florkowski and P. Bożek, Phys. Lett. B **635**, 290 (2006) [arXiv:nucl-th/0510033].

[18] S. Pratt, Phys. Rev. C **49**, 2722 (1994).

[19] J. Randrup, Acta Phys. Hung. A **22**, 69 (2005).

[20] W. N. Zhang, Y. Y. Ren and C. Y. Wong, Phys. Rev. C **74** (2006) 024908 [arXiv:hep-ph/0606113].

[21] Z. T. Yang, W. N. Zhang, L. Huo and J. B. Zhang, J. Phys. G **36**, 015113 (2009) [arXiv:0808.2413 [hep-ph]].

[22] A. Kolmogorov, Giornale dell' Instituto Italiano degli Attuari 4 (1933), 83-91.

[23] N.V. Smirnov, Mat. Sb. **6**, (1939), 3-26. [In Russian]; N.V. Smirnov, Moscow, Bull. Univ. **2**, (1939), 3-14. [In Russian].

[24] B. Tomášik, arXiv:0806.4770 [nucl-th], to be published in Copmuter Physics Communications.

[25] N.H. Kuiper, Proc. Konikl. Akad. Van Wettenschappen, Series A **63**, 38 (1960).

[26] M.A. Stephens, Biometrika **52**, 309 (1965).

[27] E. Schnedermann, J. Sollfrank and U. W. Heinz, Phys. Rev. C **48**, 2462 (1993) [arXiv:nucl-th/9307020].

[28] T. Csörgő and B. Lörstad, Phys. Rev. C **54**, 1390 (1996) [arXiv:hep-ph/9509213].

[29] F. Retiere and M. A. Lisa, Phys. Rev. C **70**, 044907 (2004) [arXiv:nucl-th/0312024].

[30] A. Andronic, P. Braun-Munzinger and J. Stachel, Nucl. Phys. A **772**, 167 (2006) [arXiv:nucl-th/0511071].

[31] M.A. Stephens, Journal of the Royal Statistical Society, Series B, **32**, No.1. (1970), 115-122.

[32] W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, Numerical Recipes (The Art of Scientific Computing), Cambridge University Press, ISBN 0 521 38330 7, 1989.

[33] C. Li-Chien, Acta Math. Sinica **6**, 55-82 (1956) [In Chinese], A Russian translation is in the *"Mathematika" collection*, Vol. **4**, 135-159, (1960).

[34] W. Peltz, I.J. Good, Journal of the Royal Statistical Society, Ser. B, **38**, 152 (1976).

[35] P.J. Kim, Journal of the American Statistical Association **64**, 1625 (1969).

[36] J.R. Brown, M.E. Harvey, in preparation, private communication.

[37] G. Marsaglia, W.W. Tsang and J. Wang, Journal of Statistical Software **8**, 1 (2003).